

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : A01N 43/16, A61K 31/715, C07H 3/06, C12P 19/26, C12N 5/04, A01H 4/00	A1	(11) International Publication Number: WO 99/07223 (43) International Publication Date: 18 February 1999 (18.02.99)
(21) International Application Number: PCT/US98/16261 (22) International Filing Date: 5 August 1998 (05.08.98) (30) Priority Data: 08/907,226 6 August 1997 (06.08.97) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 08/907,226 (CIP) Filed on 6 August 1997 (06.08.97) (71) Applicant (for all designated States except US): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 12th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): ETZLER, Marilyn, E. [US/US]; 1514 Sycamore Lane, Davis, CA 95616 (US). MURPHY, Judith, B. [US/US]; 725 North Campus Way, Davis, CA 95616 (US).	(74) Agents: SMITH, Timothy, L. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>	
(54) Title: A NOD FACTOR BINDING PROTEIN FROM LEGUME ROOTS (57) Abstract The present invention provides NBP46 polynucleotides that are useful in modulating Nod factor binding and other plant functions.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

A NOD FACTOR BINDING PROTEIN FROM LEGUME ROOTS

CROSS REFERENCE TO RELATED APPLICATIONS

This is a continuation in part of USSN 08/907,226, filed August 6, 1997, which is incorporated herein by reference.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support under Grant No. GM21882, awarded by the National Institutes of Health and under Grant No. DCB 9004967, awarded by the National Science Foundation. The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Usable nitrogen is the major limiting nutrient in crop plant growth. Plants derive most of their nutrients including nitrogen from the soil through uptake in the root system. Although most of the nitrogen in the soil is in the form of ammonium ions which is rapidly converted to usable nitrates by bacteria in the soil, the harvesting of plants results in a steady decrease of nitrogen from the soil. Unless the soil is augmented with nitrogen-containing compounds, the soil becomes depleted of usable nitrogen and only atmospheric nitrogen remains.

Legumes, unlike other higher plants, are able through a symbiotic relationship with bacteria to utilize atmospheric nitrogen in the soil. The bacteria, *Rhizobia*, infect leguminous seedlings and induce nodulation, the end result being the presence within the root system of nodules which contain the rhizobial bacteroids. Once within the root system, the bacteroids are able to "fix" atmospheric nitrogen into organic compounds the legumes can use. In exchange for the conversion of atmospheric nitrogen, the plants provide the bacteroids with carbon-containing compounds, other nutrients, and a protective environment.

Although the "fixed" nitrogen is used throughout the plant in the growth and development of its organs and tissues, much of the usable nitrogen remains within the nodules of the roots. This empirical finding has led to the practice of crop rotation wherein a non-leguminous plant, *i.e.*, corn, is grown and harvested and then the field is sown with a legume, such as alfalfa. After harvest of the legume, the remaining roots are plowed under and thus, usable nitrogen is returned to the soil for the sowing of the non-leguminous crop.

The legumes recognize the rhizobial bacteria through a lectin-carbohydrate interaction. Within the root system, the plants contain lectins that bind to specific carbohydrates found on the *Rhizobium* cell wall. This interaction is very specific; with each plant recognizing and being infected by one rhizobial strain.

In addition to their involvement in recognition of rhizobial bacteria, oligosaccharide signaling events play important roles in the regulation of plant development, defense, and other interactions of plants with the environment (Ryan, C.A. and Farmer, E.E. *Annu. Rev. Plant Physiol. Plant Mol. Bio.* 42:651-674 (1991); Cote, F. and Hahn, M.G. *Plant Mol. Biol* 26:1379-1411 (1994); Denarie, I. *et al. Annu. Rev. Biochem.* 65:503-535 (1996)). Although the structures of some of these oligosaccharides have been characterized, little is known about the plant receptors for these signals, nor the mechanism(s) by which these signals are transduced.

Previously, a root lectin, NBP46 (formerly called DB46), was isolated from young *Dolichos biflorus* root extracts. NBP46 is a 46 kDa protein that was isolated by affinity chromatography on hog gastric mucin blood group A + H substance conjugated to Sepharose (Quinn, J.M. and Etzler, M.E. *Arch. Biochem. Biophys.* 258:535-544 (1987)).

Identification and characterization of protein and the genes that encode them is important to modulation of oligosaccharide signaling in plants. For instance, a transgenic non-leguminous plant containing a factor that allows rhizobial bacteria to infect the plant and fix nitrogen would lessen the need for the addition of nitrogen-containing fertilizer to soil and preclude the necessity of crop rotation in nitrogen- depleted fields. This would lead to higher yields of crop plants in areas of the world where the soil has been overplanted and replenishment of the depleted soil with usable nitrogen. The present invention addresses these and other needs.

SUMMARY OF THE INVENTION

This invention provides for the isolation and cloning of the cDNA of *NBP46* (SEQ ID NO:1), which encodes NBP46, a Nod factor binding lectin. Nod factors are carbohydrates on the surface of *Rhizobium* which bind to lectins on the surface of leguminous plant organs and can initiate nodulation of the root system by the plants. The *NBP46* gene encodes a polypeptide of between 50 and 560 amino acids, more preferably 462 amino acids (SEQ ID NO:2).

In a preferred embodiment, the *NBP46* coding sequence is operably linked to a plant specific promoter, more preferably a root specific promoter, such as the *NBP46* promoter (SEQ ID NO:3).

In another embodiment, an expression cassette comprising the *NBP46* gene is introduced into a transgenic plant. In a preferred embodiment, the expression of NBP46 by the transgenic plant confers to the plant the ability to bind to rhizobial bacteria and utilize atmospheric nitrogen. In a particularly preferred embodiment, the expression of NBP46 confers to the plant the ability to catalyze the hydrolysis of the phosphoanhydride bonds of di- and tri-phosphates, leading to greater availability of nutrients to the plant.

In a further embodiment of the instant invention, methods of modulating the rhizobial interactions and in the phosphatase activity in plants by the introduction of an expression cassette comprising *NBP46* are disclosed.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 indicates the inhibition of binding of ^{125}I -NBP46 to HBG A + H-Sepharose®.

In Figure 1 A, the legend is as follows: HBG A + H (■); human ovarian cyst blood group A substance (◆); human ovarian cyst blood group H substance (▼); de-N-acetylated HBG A + H (●).

In Figure 1 B, the legend is as follows: *Bradyrhizobium japonicum* USDA 110 Nod factor (■); β -O-methyl galactose β (1-3) N-acetyl-D-glucosamine (O); methyl α -N-acetyl-D-glucosamine (●); methyl β -N-acetyl-D-glucosamine (◆); dimer (Δ), trimer (\square), and tetramer (O) of β (1-4) N-acetyl-D-glucosamine.

Figure 2 shows the effect of carbohydrate ligands on phosphatase activity of NBP46, NBP46 (201 ng/ml) was preincubated for 1 hour in the presence of various

concentrations of *B. japonicum* USDA110 Nod factor (■), *R. sp.* NGP,234(Ac) Nod factor (▼), *R. sp.* NGR234(S) Nod factor (▲), *R. meliloti* Nod factor (●), or *cis*-vaccenic acid (◆) and then assayed for phosphatase activity using a final concentration of 3 mM Mg-ADP.

Figure 3 shows inhibition of binding of ¹²⁵I-NBP46 to chitin. Various concentrations of mono- and oligosaccharides were combined with 109 ng ¹²⁵I-NBP46 and 250 μg of chitin in a total volume of 100 μl. *B. japonicum* USDA110 Nod factor (■); *R. sp.* NGR234(NGR_A) Nod factor (▲); *R. sp.* NGR234(NGR_B) Nod factor (▼); *R. meliloti* Nod factor (●), *N*-acetylglucosamine (□), chitin disaccharide (▼); chitin tetrasaccharide (Δ); chitin pentasaccharide (◆), chitin hexasaccharide (O).

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

The phrase "isolated nucleic acid molecule" or "isolated protein" refers to a nucleic acid or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *NBP46* gene is separated from open reading frames which flank the gene and encode a protein other than *NBP46*. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most

environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R₁ generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

The phrase "rhizobial binding" refers to the binding between rhizobial bacteria and plant cells. Typically, enhanced binding leads to infection by rhizobial bacteria of the roots of plants. This in turn leads to nodule formation in the roots. For example, a non-leguminous transgenic plant comprising a polynucleotide of this invention and expressing its corresponding polypeptide in the roots of the plant would bind to Nod factors of rhizobial bacteria allowing the plant to become infected by the rhizobial bacteria

and allowing the plant to reduce the atmospheric nitrogen contained in the soil and using it as a nutrient.

The phrase "operably linked" refers to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "polynucleotide," "polynucleotide sequence" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular *NBP46* nucleic acid sequence of this invention also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.*, 1992; Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

A "*NBP46* polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of about 100 to about 2000 nucleotides, sometimes from about 1400 to about 1500 nucleotides, which hybridizes to SEQ ID NO:1 under stringent conditions (as defined below), or which encodes a *NBP46* polypeptide.

The term "sexual reproduction" refers to the fusion of gametes to produce seed by pollination. A "sexual cross" is pollination of one plant by another. "Selfing" is the production of seed by self-pollinization, *i.e.*, pollen and ovule are from the same plant.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *NBP46* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "NBP46 nucleic acid".

5 In addition, the term specifically includes those sequences substantially identical (determined as described below) with an *NBP46* polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type *NBP46* polypeptides or retain the function of the *NBP46* polypeptide (e.g., resulting from conservative substitutions of amino acids in the *NBP46* polypeptide). In addition, variants can be those
10 that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide
15 sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in
20 reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be
25 adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a
30 conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer*

Applic. Biol. Sci. 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA)..

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to sequences or subsequences that have at least 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has substantial sequence or subsequence complementarity when the test sequence has substantial identity to a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP
5 uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster
10 of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence
15 comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent
20 sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence,
25 which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be
30 increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-

scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments
5 (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which
10 a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic
15 acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the
20 codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes
25 every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

30 As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino

acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

5 The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 10 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, *Proteins* (1984)).

15 An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative
20 substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

 The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under
25 stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

 The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and
30 will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic*

Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30 °C below the T_m . The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising *NBP46* nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as

a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

The phrase "transgenic plant" refers to a plant into which heterologous polynucleotides have been introduced by any means other than sexual cross or selfing.

5 Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the heterologous polynucleotides is referred to here as an R₁ generation transgenic plant. Transgenic plants may also arise from sexual cross or by selfing of transgenic plants into which heterologous
10 polynucleotides have been introduced.

II. Introduction

The present invention provides polynucleotides referred to here as *NBP46* polynucleotides, as exemplified by SEQ ID NO:1. Polypeptides encoded by the genes of
15 the invention are lectins involved in binding a variety of carbohydrates. In addition, polypeptides function as an enzyme, catalyzing the dephosphorylation of nucleotide di- and triphosphates. As explained below, the nucleic acid sequences of the invention code for a Nod factor binding lectin naturally expressed in the root tissue of leguminous plants.

The polypeptides of the invention are also involved oligosaccharide
20 signaling events that play important roles in the regulation of plant development, defense, and other interactions of plants with the environment. Although the structures of some of these oligosaccharides have been characterized in the prior art, little is known about the plant receptors for these signals, nor the mechanism(s) by which these signals are transduced. The results presented below show that polypeptides of the invention serve as
25 receptors in oligosaccharide signaling.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase,
30 restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook, *et al.*

III. Isolation Of Nucleic Acid Sequences From Plants

The isolation of sequences from the genes of the invention may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the nucleic acid and peptide sequences disclosed herein can be used to identify the desired gene in a cDNA or genomic DNA library from a desired leguminous plant species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, *e.g.*, using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a library of tissue-specific cDNAs, mRNA is isolated from tissues and a cDNA library which contains the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned gene such as the polynucleotides disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes known to those of skill.

Appropriate primers and probes for identifying *NBP46* genes from *Dolichos biflorus* or transgenic plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Appropriate degenerate primers for this invention include, for instance: a 5' PCR primer [5'-TA(T/C)GCNGTNAT(T/C)TT(T/C)GATGC-3'] (SEQ ID NO:4) and a 3' PCR primer [5'-AT(A/G)TT(A/G)TA(T/A/G)AT(G/A)CCNGG-3'] (SEQ ID NO:5) where N denotes all nucleotides. The amplification conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μ M dATP, 200 μ M dCTP,

200 μ M dGTP, 200 μ M dTTP, 0.4 μ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.

Using the above primers, a partial coding sequence will be obtained. There are many techniques known to those of skill to determine and isolate the complete coding sequence. These methods include using the PCR amplified subsequence to probe a cDNA library for longer sequences.

A preferred method is RACE (Frohman, *et. al.*, *Proc. Nat'l. Acad. Sci. USA* 85:8998 (1988)). Briefly, this technique involves using PCR to amplify a DNA sequence using a random 5' primer and a defined 3' primer, *e.g.*, (SEQ ID NO:6) (5' RACE) or a random 3' primer and a defined 5' primer, *e.g.*, (SEQ ID NO:7) (3' RACE). The amplified sequence is then subcloned into a vector where it is then sequenced using standard techniques. Kits to perform RACE are commercially available (*e.g.* 5' RACE System, GIBCO BRL, Grand Island, New York, USA). In this manner, the entire NBP46 coding sequence of about 1600 bp can be obtained (SEQ ID NO:1). The invention also provides genomic sequence of the NBP46 (SEQ ID NO:3).

Alternatively, primers can be selected and synthesized by those of skill from the cDNA sequence disclosed in SEQ ID NOs:1 and 3.

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, *e.g.*, Carruthers, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

IV. Use Of Nucleic Acids Of The Invention To Modulate Gene Expression

The polynucleotides of the invention can be used to enhance expression (*i.e.*, increase expression of an endogenous gene or provide NBP46 expression in a plant that does not normally express NBP46) of genes of the invention and thereby enhance infection of transgenic plants by rhizobial bacteria, increase the level of nutrients taken up by the plants, and affect the growth and development of transgenic plants. Alternatively,

enhanced expression can be used to modulate oligosaccharide signaling in the plant. This can be accomplished by the overexpression of NBP46 polypeptides in the tissues of transgenic plants.

The heterologous *NBP46* polynucleotides do not have to code for exact
5 copies of the NBP46 proteins exemplified herein. Modified NBP46 polypeptide chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described for instance, in Sambrook *et al.*, *supra*.

Hydroxylamine can also be used to introduce single base mutations into the coding region of the gene (Sikorski, *et al.*, *Meth. Enzymol.* 194: 302-318 (1991)). For example, the
10 chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Alternatively, the nucleic acid sequences of the invention can be used to inhibit expression of an endogenous gene. One of skill will recognize that a number of
15 methods can be used to inactivate or suppress NBP46 activity or gene expression. The control of the expression can be achieved by introducing mutations into the gene or using recombinant DNA techniques. These techniques are generally well known to one of skill and are discussed briefly below.

Methods for introducing a genetic mutations into a plant genes are well
20 known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used. Desired mutants are
25 selected by assaying for increased seed mass, oil content and other properties.

Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. *NBP46* mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of
30 *NBP46* mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for increased seed mass, oil content and other properties.

The isolated sequences prepared as described herein, can also be used in a number of techniques to suppress endogenous *NBP46* gene expression. A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy et al., *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt et al., U.S. Patent No. 4,801,340.

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous *NBP46* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 1700 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *NBP46* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

A. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising, *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length NBP46 protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transgenic plant, *i.e.*, a root-specific promoter.

Promoters can be identified by analyzing the 5' sequences of a genomic clone in which naturally occurring Nod factor binding protein-specific genes, *i.e.*, *NBP46*, can be found. At the 5' end of the coding sequence, nucleotide sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing, *et al.*, in *GENETIC ENGINEERING IN PLANTS*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (*see, e.g.*, Jordano, *et al.*, *Plant Cell* 1:855-866 (1989); Bustos, *et al.*, *Plant Cell* 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7:4035-4044 (1988); Meier, *et al.*, *Plant Cell* 3:309-316 (1991); and Zhang, *et al.*, *Plant Physiology* 110:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus

(CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter may direct expression of the polynucleotide of the instant invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues, such as roots, fruit, seeds, or flowers. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (e.g., promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

B. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of a desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of a plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly into plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, *et al.*, *EMBO J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, *et al.*, *Proc. Nat'l. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein, *et al.*, *Nature* 327:70-73 (1987).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, *et al.*, *Science* 233:496-498 (1984), and Fraley, *et al.*, *Proc. Nat'l. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, *PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE*, pp. 124-176, Macmillian Publishing Company, New York (1983); and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

To determine the presence of a reduction or increase of NBP46 activity, a variety of assays can be used including enzymatic, immunochemical, electrophoretic detection assays (either with staining or western blotting), or complex carbohydrate binding assays.

In a preferred embodiment, a competitive solid phase assay is used to measure NBP46 activity (Etzler, M.E., *Glycoconj. J.* 11:395 (1994)). This assay measures the ability of various ligands to inhibit the binding of labeled NBP46 protein to pronase-digested hog gastric mucin blood group A + H substance (HBG A + H) conjugated to Sepharose® (Quinn, J.M. & Etzler, M.E., *Arch. Biochem. Biophys.* 258:535 (1987)).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Asparagus*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Glycine*, *Hordeum*, *Lactuca*, *Lycopersicon*, *Malus*, *Manihot*,
5 *Nicotiana*, *Oryza*, *Persea*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna*, and *Zea*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be
10 used, depending upon the species to be crossed.

Effects of gene manipulation can be observed by northern blots of the mRNA isolated from the tissues of interest. Typically, if the amount of mRNA has increased, it can be assumed that the gene is being expressed at a greater rate than before. Other methods of measuring NBP46 expression would be by measuring the rhizobial
15 infection of the transgenic plants. Alternatively, the ability of the plant to reduce atmospheric nitrogen could be assessed. In addition, levels of NBP46 could be measured immunochemically, *i.e.*, ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art.

20 V. Examples

The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Characterization and cloning of NBP46

25 Carbohydrate binding and characterization of NBP46

It has been previously demonstrated that NBP46 is a 46 kDa protein that can be isolated from young *Dolichos biflorus* root extracts by affinity chromatography on hog gastric mucin blood group A + H substance (HBG A + H) conjugated to Sepharose® (Quinn, J.M. & Etzler, M.E., *Arch. Biochem. Biophys.* 258:535 (1987)). The monomeric
30 nature of NBP46 in solution precluded the use of conventional precipitin or agglutination assays in determining the carbohydrate binding specificity of this lectin. Therefore a

complex carbohydrate binding assay was employed (Etzler, M.E., *Glycoconj. J.* 11:395 (1994)).

As shown in Figure 1, various concentrations of blood group substances (A) and oligosaccharides (B) were combined with 109 ng 125 I-NBP46 (isolated as described in Quinn, J.M. & Etzler, M.E., *Arch. Biochem. Biophys.* 258:535 (1987)) and a pronase digest of HBG A + H-Sepharose® (final concentration 1%) in a volume of 100 μ L of 5 mM MOPS, pH 7.2, containing 0.025% Tween-20® and 0.01% NaN₃. Hog blood group A + H substance was isolated from hog gastric mucin (Etzler, M.E., *Glycoconj. J.* 11:395 (1994)) and de-N-acetylated as described in Etzler, M.E., *et al.*, *Arch. Biochem. Biophys.* 141:588 (1970). After incubation at room temperature overnight, binding was measured as previously described (Etzler, M.E., *Glycoconj. J.* 11:395 (1994)). Although the binding of the NBP46 to this resin was inhibited by free HBG A + H (Figure 1A), no significant inhibition was obtained with up to 50 mM concentrations of any of the monosaccharides present in the blood group substance, including N-acetyl-D-galactosamine and L-fucose, the immunodominant sugars of the blood type A and H determinants, respectively (Watkins, W.M., *Science* 152:172 (1966); and Lloyd, K.O., *et al.*, *Proc. Nat'l. Acad. Sci. USA* 61:1470 (1968)). Individual human ovarian cyst blood group A and H substances (provided by Elvin A. Kabat, Columbia University) were equal to one another in inhibitory capacity but much weaker than HBG A+H (Figure 1A). De-N-acetylation of the blood type A determinant did not alter the ability of the HBG A + H to inhibit the binding of NBP46 (Figure 1A).

These results indicated that the binding of NBP46 to the above blood group substances was due to its recognition of some portion of the oligosaccharide chains other than the blood type A and H determinants and that its carbohydrate binding site accommodated more than a simple sugar. The carbohydrate specificity of NBP46 thus differs from the blood type A specific seed lectin from *Dolichos biflorus*, which recognizes the α N-acetyl-D-galactosamine residues which are at the nonreducing ends of the oligosaccharide chains of blood group A substance (Etzler, M.E., *et al.*, *Biochemistry* 9:869 (1970)).

A variety of oligosaccharides were tested in an attempt to obtain more information on the carbohydrate specificity of NBP46 (Figure 1B). The strongest inhibition was obtained with the purified Nod factor from *Bradyrhizobium japonicum*

USDA 110, a bacterial rhizobial strain that nodulates soybean and can also nodulate *Dolichos biflorus*. The Nod factor was isolated as described in Sanjuan, J., *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:8789 (1992). The Nod factor from *Bradyrhizobium japonicum* USDA 110 is composed of a β 1-4 *N*-acetyl-D-glucosamine pentasaccharide backbone, modified by a 2-*O*-methyl α -L-fucose on C-6 of the sugar at the reducing end and the substitution of the acetyl group on the sugar at the nonreducing end with a C_{18:1} fatty acyl chain (Sanjuan, J., *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:8789 (1992)). Thus, NBP46 can be characterized as a Nod factor binding lectin.

Phosphohydrolase Activity of NBP46

A search of protein and nucleotide data bases using the NCB1 *BLASTP* and *BLASTN* programs (Altschul, S.F., *et al.*, *J. Mol. Biol.* 215:403 (1990)) showed no significant similarities between NBP46 to the amino acid or cDNA sequences of any other plant or animal lectin yet described. It did, however, show 65.6 and 47.6% amino acid identity and 70.7 and 58.7% nucleotide identity with the sequences of a pea nucleotide triphosphatase (Hsieh, H.-L., *et al.*, *Plant Mol. Biol.* 30:135 (1996), GenBank Accession No. Z32743) and an apyrase isolated from potato tubers (Handa, M., *et al.*, *Biochem. Biophys. Res. Comm.* 218:916 (1996)). Thus, the pea triphosphatase gene could also be used in the methods of the invention. Considerably less, but significant, similarity was found with the sequences of several other animal and yeast phosphohydrolases. Of particular interest in this comparison was the presence in all of these sequences of four motifs (designated by the bold letters in SEQ ID NO: 2) identified as conserved regions among a variety of plant and animal apyrases (Handa, M., *et al.*, *Biochem. Biophys. Res. Comm.* 218:916 (1996)).

The sequence similarities found between NBP46 and the above enzymes prompted the testing of NBP46 for phosphohydrolase activity. The reactions were conducted in 300 μ L of 60 mM MOPS, pH 6.8, containing 1 mM MgCl₂ in a microtiter plate using a multichannel pipette. At various time points up to 4 minutes, 30 or 60 μ L aliquots were removed and assayed for inorganic phosphate by a photometric microtiter assay (Drueckes, P., *et al.*, *Anal. Biochem.* 230:173 (1995)). Conditions were chosen so that less than 10% of the total substrate was converted to product, and the initial velocity (*v*) was determined from the above rate measurements. The *K_m* of NBP46 for Mg-ADP was found to be 615 μ M.

NBP46 catalyzed the hydrolysis of phosphate from both ATP and ADP (Figure 2) but showed no activity with AMP, pyrophosphate or glucose-6-phosphate. It also had a broad specificity for nucleotide triphosphates, including GTP, CTP and UTP. This substrate specificity has been found to be characteristic of the apyrase category of phosphohydrolases (EC 3.6.1.5). Preincubation of NBP46 with 10 µg/mL of HBG A + H (which results in 46% inhibition of carbohydrate binding activity) resulted in an increase in the V_{\max} of NBP46. No increase in phosphatase activity was observed upon preincubation of NBP46 with human blood group H substance at a concentration that shows no inhibition in the carbohydrate binding assay described above (Figure 2). The V_{\max} of NBP46 was also increased in the presence of low concentrations (1 to 5 micromolar) of Nod factors, with lower concentrations required for the Nod factors produced by rhizobia that nodulate the plant than for the *R. meliloti* Nod factor. These results suggest that there is interaction between the carbohydrate binding and phosphatase sites of NBP46.

Isolation and Characterization of NBP46 cDNA and Encoded Protein

Two consensus *N*-glycosylation sites are present in the sequence of the mature protein at residues 111 and 276. Work in progress in our laboratory has established that NBP46 is indeed glycosylated at at least one of these sites. It should be noted, however, that we do not yet know whether other posttranslational modifications of this protein may occur, such as the COOH-terminal proteolysis that modifies two other lectins from this plant (Etzler, M.E. *Biochemistry* 33:9778-9783 (1994); Schnell, D.T. *et al. Arch. Biochem. Biophys.* 310:229-235 (1994)). A search of protein and nucleotide data bases using the NCB1 *TBLASTN* and *BLASTN* programs (Altschul, S.F. *et al. J. Mol. Biol.* 215:403-410 (1990)) showed no significant similarities of NBP46 to the amino acid or cDNA sequences of any other plant or animal lectin yet described. It did, however, show 65.6 and 47.6% amino acid identity and 70.7 and 58.7% nucleotide identity with the sequences of a pea nucleotide triphosphatase (Hsieh, H-L. *et al. Plant Mol. Biol.* 30:135-147 (1996)) and an apyrase isolated from potato tubers (Handa, M. and Guidotti, G. *Biochem. Biophys. Res. Comm.* 218: 916-923 (1996)), respectively. Considerably less, but significant, similarity was also found with the sequences of several other animal and yeast phosphohydrolases. Of particular interest in this comparison is the presence in all of these sequences of four motifs (designated by the boxes in SEQ ID NO: 2) identified as

conserved regions among a variety of plant and animal apyrases (Handa, M. and Guidotti, G. *Biochem. Biophys. Res. Comm.* 218: 916-923 (1996)).

The sequence similarities found among NBP46 and the above enzymes prompted us to test NBP46 for phosphohydrolase activity. NBP46 catalyzes the hydrolysis of phosphate from both ATP and ADP but showed no activity with AMP, pyrophosphate or glucose-6-phosphate. The K_m of NBP46 for Mg^{++} -ADP is 615 μ M. The lectin has a broad specificity for nucleotide triphosphates, including GTP, CTP and UTP (data not shown). This substrate specificity is characteristic of the apyrase category of phosphohydrolases (EC 3.6.1.5). Preincubation of NBP46 with ligands that are recognized by its carbohydrate binding site results in an increase in the V_{max} of this enzyme. Low micromolar concentrations of the above Nod factors stimulate this increase in activity, with lower concentrations required for the Nod factors produced by rhizobia that modulate the plant than for the *R. meliloti* Nod factor (Figure 2). Such an increase in enzyme activity is also obtained with low millimolar concentrations of the chitin oligosaccharides and *N*-acetylglucosamine, but not with *N*-acetylgalactosamine (data not shown). These results suggest that there is interaction between the carbohydrate binding and phosphatase sites of NBP46. Whether this interaction represents a direct stimulation of the enzyme activity or perhaps a stabilization of the enzyme under the assay conditions remains to be determined.

NBP46 binds to chitin and other carbohydrates

NBP46 also binds to chitin, a polymer of $\beta(1-4)$ linked *N*-acetyl-D-glucose residues; this binding is saturable with a B_{max} of 28 nmoles of NBP46/gram of chitin and a K_d of 48 nM. Using chitin as a solid phase, a competitive binding assay was utilized to examine the carbohydrate specificity of this protein (Figure 3). Inhibition of binding was obtained with high concentrations of *N*-acetyl-D-glucosamine but not with similar concentrations of *N*-acetyl-D-galactosamine, the C4 epimer of this sugar, nor with other common monosaccharides. The chitin disaccharide gave approximately ten-fold better inhibition than the monosaccharide, whereas the chitin penta- and hexasaccharides were slightly better inhibitors than the disaccharide. No inhibition was obtained with the de-*N*-acetylated chitin oligosaccharides; however, when tested in the millimolar range of concentrations, several of these oligosaccharides precipitated the lectin even under highly buffered conditions. Whether this precipitation is specific or nonspecific is under investigation.

Of all the oligosaccharides tested, the best inhibition was obtained with the Nod factor isolated from *Bradyrhizobium japonicum* USDA110 (Figure 3), a rhizobial strain that nodulates *Dolichos biflorus*. The chitolipo-saccharidic Nod factors have been identified as the signals produced by rhizobia that initiate the nodulation of legumes (Denarie, I. *et al. Annu. Rev. Biochem.* 65:503-535 (1996)). The *B. japonicum* USDA110 Nod factor consists of a chitin pentasaccharide backbone, modified by a 2-*O*-methyl α -L-fucose on C-6 of the sugar at the reducing end and the substitution of the acetyl group on the sugar at the nonreducing end with *cis*-vaccenic acid (Sanjuan, J. *et al. Proc. Natl. Acad. Sci. USA*, 89:8789-8793 (1992); Carlson, R.W. *et al. J. Biol. Chem.* 268:18372-18-81 (1993)). The higher relative affinity of NBP46 for the intact Nod factor than for the chitin pentasaccharide backbone alone indicates that the modifications of this backbone contribute to the recognition of the Nod factor by the lectin. No significant inhibition of NBP46 binding to chitin was obtained with *cis*-vaccenic acid when tested at concentrations up to 1.2 mM nor with L-fucose at concentrations up to 50 mM.

Two Nod factors from *Rhizobium* sp. NGR234, another strain that nodulates *Dolichos biflorus*, were also able to inhibit the binding of NBP46 to chitin. These Nod factors differ from the USDA110 Nod factor in that they have a sulfate on C-3 (NodNGR_A) or an acetate on C-4 (NodNGR_B) of the 2-*O*-methylfucose; they are also methylated on the amino group and partially carbamoylated at C-3, C-4 or C-6 of the sugar at the nonreducing end (Price, N.P.J. *et al. Carbohydr. Res.* 289:115-136 (1993)). The Nod factor from *Rhizobium meliloti*, a strain that does not modulate *Dolichos biflorus*, gave the weakest inhibition when tested at equivalent concentrations (Figure 3). This Nod factor differs from the USDA110 Nod factor in that it has a chitin tetrasaccharide backbone, contains a sulfate instead of a fucose at the reducing end and is acetylated at C-6 of the sugar at the nonreducing end (Lerouge, P. *Nature* 344:781-784 (1990)).

Although the differences in relative affinity of NBP46 for the above Nod factors indicate a small preference of the lectin for Nod factors produced by rhizobia that modulate the plant, it must be pointed out that both the *B. japonicum* USDA110 and *R. sp.* NGR234 strains are only weak nodulators of *Dolichos biflorus*, and the nodules formed with the former strain do not fix nitrogen. Nod factors from rhizobial strains that are strong nodulators of this plant have not yet been purified or characterized.

Antiserum raised against NBP46 inhibits nodulation

Confocal immunofluorescence microscopy of whole mounts of 7-day old *Dolichos biflorus* roots that had been fixed prior to staining showed that NBP46 is present on the surfaces of the newly emerging and young root hairs. Treatment of young roots of this plant with antiserum to the lectin inhibited the ability of these roots to be nodulated by rhizobia (Table 1). Although it is possible that such inhibition could be due to steric hindrance of adjacent sites, these results, coupled with the above finding that NBP46 is a Nod factor binding protein, suggest that this root lectin may play a role in rhizobium-legume symbiosis either as a host/strain specific receptor or perhaps as a second, less stringent receptor postulated for this process (Ardourel, M. *et al. Plant Cell* 6:1357-1374 (1994)). Previous attempts to implicate lectins in this symbiosis have been focused on the legume seed lectins (Diaz, C.L. *et al. Nature* 338:579-581 (1989); Hirsch, A.M. *et al. Symbiosis* 19:155-173 (1995)), which have not been reported to bind Nod factors. It is also possible that NBP46 may function in the recognition of endogenous Nod-factor like signals that have been proposed to play a role in the regulation of plant growth and organogenesis (Etzler, M.E. *Biochemistry* 33:9778-9783 (1994)).

**Table 1. Effect of anti-NBP46-serum
on nodulation of *D. biflorus* roots**

Treatment	Average number of nodules (\pm S.E.)	
	Treated region of root	Region of root emerged after treatment
Untreated	3.6 \pm 0.5	2.2 \pm 0.2
Preimmunization serum	3.4 \pm 0.5	1.6 \pm 0.2
Anti-NBP46-serum	0.6 \pm 0.2	1.4 \pm 0.2

The roots of 2 sets of 10 3-day old *Dolichos biflorus* plants were immersed for 1 hour in 1/100 dilutions of preimmunization serum or anti-NBP46-serum, washed and transferred to growth pouches. A third set of 10 pts was put directly in growth pouches. Half of each set of plants was inoculated with *Bradyrhizobium* sp. 24A10. After 3 weeks the number of nodules in the treated region as well as in the region of root that emerged after treatment were recorded. No nodules were observed on the roots that had not been inoculated with rhizobia.

DISCUSSION

The low concentrations (10^{-12} of Nod factor that have been found to induce physiological responses in legumes (Denarie, I. *et al. Annu. Rev. Biochem.* 65:503-535 (1996)) predict that Nod factor receptors have high affinity for their ligands. Indeed, high affinity binding sites for Nod factors have been found on particulate fractions from roots of the legume, *Medicago truncatula* (Niebel, A. *Mol. Plant-Microbe Interact.* 10:132-134 (1997)). Although the inhibition data show the relative affinities of NBP46 for its ligands, they do not enable the determination of the absolute affinities of this lectin for the Nod factors. The concentrations of Nod factors required for the stimulation of increased phosphatase activity suggest that the K_d 's may be in the high nanomolar to low micromolar range. It should be noted, however, that NBP46 is primarily a monomer in solution (Quinn, J.M. and Etzler, M.E. *Arch. Biochem. Biophys.* 258:535-544 (1987)); as established with antibodies (Hornick, C.L. and Karush, F. *Immunochem.* 9:325-340 (1972)), the multivalence that would occur when this lectin is associated with the cell surface would increase its apparent affinity for multivalent ligands such as Nod factor micelles or Nod factor on the surface of rhizobia by several orders of magnitude.

The presence of both carbohydrate binding activity and apyrase activity on NBP46 and the apparent interaction of these sites suggest that, upon binding its carbohydrate ligand, NBP46 may play a role in activating downstream events either directly by signal transduction or indirectly, perhaps by serving as a motor for transport of the carbohydrate. In this context, it is of interest that the human CD39 lymphoid cell activation antigen, one of the apyrases found to have some sequence similarity to NBP46, is thought to be involved in the regulation of B cell adhesion (Kansas, G.S. *et al. J. Immunol.* 146:2235-2244 (1991)). Although these other apyrases have not been tested for lectin activity, it is possible that such dual activities of these proteins may have been conserved throughout evolution.

The unique amino acid sequence, carbohydrate specificity and apyrase activity of NBP46 distinguish this lectin from the conventional lectins found in abundance in the seeds of legumes (Sharon, N. and Lis, H. *FASEB J.* 4:3198-3208 (1990)). The possibility that other such plant lectin/enzymes exist is suggested by the recent finding of a CDNA from *Arabidopsis thaliana* that encodes a receptor-like serine/threonine kinase as well as a legume seed lectin-like domain (Herve, C. *et al. J. Mol. Biol.* 258:778-788

(1996)). NBP46 may thus be on one of many multifunctional carbohydrate binding proteins that may function in plant oligosaccharide signaling events. A variety of transgenic experiments are underway to elaborate its role in such processes.

5 METHODS

Preparation of NBP46. NBP46 was extracted from the roots of 7-day old *Dolichos biflorus* plants and isolated by affinity chromatography on hog blood group A + H -Sephrose as previously described (Quinn, J.M. and Etzler, M.E. *Arch. Biochem. Biophys.* 258:535-544 (1987)), followed by ion exchange chromatography. It was
10 iodinated using the iodine monochloride procedure as previously described (Etzler, M.E. *Glycoconj. J.* 11:395-399 (1994)), which gave a specific activity of approximately 500 x 10⁶ cpm/mg protein.

Carbohydrate binding assays. Solid phase binding assays were conducted using iodinated NBP46 and purified shrimp chitin powder (Sigma Chemical Company, St.
15 Louis, MO), which was *N*-acetylated prior to use with 15 mM acetic anhydride in 0.5 M NaHCO₃ for one hour at room temperature. The assays were conducted in a final volume of 100 µl of 10 mM MOPS buffer, pH 7.2, containing 0.02% Tween-20 and 0.01% NaN₃. After incubation at room temperature for two hours, binding was measured as previously described (Etzler, M.E. *Glycoconj. J.* 11:395-399 (1994)).

20 *Bradyrhizobium japonicum* USDA110 Nod factor was isolated as previously described (Sanjuan, J. *et al. Proc. Natl. Acad. Sci. USA*, 89:8789-8793 (1992)). The Nod factors from *Rhizobium meliloti* and *Rhizobium* sp. NGR234 were graciously provided by Dr. Jean Denarie, CNRS-INRA, Toulouse, France. Monosaccharides and the chitin disaccharide were purchased from Sigma Chemical Co., St. Louis, MO, the other
25 chitin oligosaccharides were obtained from Seikagaku Corp., Tokyo, Japan.

Cloning of NBP46 cDNA. Total RNA was isolated (Taylor, B. and Powell, A. *Focus* 4:4-6 (1982)) from the roots of 1-day-old *D. biflorus* plants and reverse transcribed using M-NMV reverse transcriptase and random hexanucleotide primers (Tabor S. RNA-dependent DNA polymerases. In *Current Protocols in Molecular Biology*, F.M. Ausubel, F.M., *et al.*, Eds., John Wiley & Sons, Inc., Vol. 1, pp. 3.7.1-3.7.3 (1987)). This
30 cDNA was used as a template in a PCR reaction employing *Taq* polymerase and degenerate sense and antisense primers corresponding to amino acids 6-12 and 244-249 in

SEQ ID NO:2. The PCR was performed in an automated thermal cycler for 35 cycles of 94°C for 2 min, 37°C for 2 min, and 72°C for 2 min. The predominant 727 bp fragment was isolated on a 1.2% agarose gel, cloned into the pCRII vector (InVitrogen) and sequenced (Sanger, F. *et al. Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977)). Gene specific primers were used in 5' and 3' RACE reactions (Frohman, M.A. *Proc. Natl. Acad. Sci USA* 85:8998-9002 (1988)); the products were cloned into the pCRII vector and sequenced. The full length (1527 bp) cDNA was assembled by ligating the two RACE products together using an int *Sac*I site. The sequences of the overlapping regions of the 5' and 3' RACE products and the original PCR fragment were identical.

Phosphatase assays. N-BP46 (201 ng/ml) was incubated at 25°C in the presence of various concentrations of substrate in a final volume of 100 µl of 60 mM MOPS, pH 6.8, containing 1 mM MgCl₂. The reactions were conducted in a microtiter plate using a multichannel pipette. At various time points, 30 µl aliquots were removed and assayed for inorganic phosphate by a photometric microtiter assay (Drueckes, P. *et al. Anal. Biochem.* 230:173 (1995)), modified by using four parts ammonium molybdate reagent to one part 10% ascorbate for the reagent mixture. Conditions were chosen so that less than 10% of the total substrate was converted to product.

Immunofluorescence microscopy. Roots from 7-day old *Dolichos biflorus* plants were fixed for 45 minutes at 4°C in 0.01 M phosphate buffer, pH 7.2, containing 0, 15 M NaCl and 0.3% paraformaldehyde. After washing, the roots were treated for 20 minutes with a 1/250 dilution of preimmunization serum or antiserum prepared against recombinant NBP46. After washing, the roots were treated for 20 minutes with fluorescein- labeled goat anti-rabbit IgG (Sigma Chemical Co., St. Louis, MO), washed and examined with a Leica TCS NT confocal microscope using a 488 nm laser excitation line and a 560 barrier filter. Confocal images were reconstructed with Imagespace software.

Nodulation. *Dolichos biflorus* seeds were sterilized by shaking for 15 minutes in 70% ethanol, followed by 15 minutes in 3% hydrogen peroxide. After extensive washing with sterile H₂O, the seeds were germinated and grown in sterile growth pouches. At 3 days, the roots were inoculated with 100 µl of *B. sp.* 24A10 (1 x 10⁷ cells/ml). The number of nodules per root was determined after 3 weeks. Antiserum and preimmunization serum used to treat the roots were sterilized by filtration through a 0.45 µm filter.

Example 2: Isolation of NBP46 from other species

NBP46 nucleic acids have also been isolated from *Medicago sativa* (SEQ ID NO:8 and 9) and *Lotus japonicus* (SEQ ID NO:10 and 11). These nucleic acids were obtained by RT-PCR as follows. Messenger RNA was obtained from the roots of both species and reverse transcribed using oligo-dT primers. Degenerate PCR primers were designed to conserved sequences of the *D. biflorus* NBP46 disclosed here and the *Pisum sativa* nucleotide triphosphatase gene described by Hsieh, H.-L., *et al.*, *Plant Mol. Biol.* 30:135(1996). These were used to generate internal 850 bp fragments from both *Medicago sativa* and *Lotus japonicus*. cDNA species-specific primers then designed for both 5' and 3' RACE. Full length clones were obtained using primers designed to the 5' and 3' ends of the RACE products. Duplicate clones from each species were obtained in separate PCR reactions and sequenced in their entirety in both directions.

Example 3: Isolation of DBX from *D. biflorus*

A second gene also involved in oligosaccharide signaling has been isolated from *D. Biflorus* (SEQ ID NO:12 and 13).

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

WHAT IS CLAIMED IS:

- 1 1. An isolated nucleic acid molecule comprising a *NBP46*
2 polynucleotide sequence, which polynucleotide sequence specifically hybridizes to SEQ
3 ID NO:1 under stringent conditions.
- 1 2. The isolated nucleic acid molecule of claim 1, wherein the *NBP46*
2 polynucleotide is between about 100 nucleotides and about 1600 nucleotides in length.
- 1 3. The isolated nucleic acid molecule of claim 1, wherein the *NBP46*
2 polynucleotide is SEQ ID NO: 1.
- 1 4. The isolated nucleic acid molecule of claim 1, further comprising a
2 plant promoter operably linked to the *NBP46* polynucleotide.
- 1 5. The isolated nucleic acid molecule of claim 4, wherein the plant
2 promoter is a root specific promoter.
- 1 6. The isolated nucleic acid molecule of claim 1, wherein the *NBP46*
2 polynucleotide encodes a NBP46 polypeptide of between about 50 and about 460 amino
3 acids.
- 1 7. The isolated nucleic acid molecule of claim 6, wherein the NBP46
2 polypeptide has an amino acid sequence as shown in SEQ ID NO:2.
- 1 8. An isolated nucleic acid molecule comprising a *NBP46*
2 polynucleotide sequence, which polynucleotide sequence encodes a NBP46 polypeptide of
3 between about 50 and about 210 amino acids.
- 1 9. The isolated nucleic acid molecule of claim 8, wherein the NBP46
2 polypeptide has an amino acid sequence as shown in SEQ ID NO:2.

1 10. A transgenic plant comprising an expression cassette containing a
2 plant promoter operably linked to a heterologous *NBP46* polynucleotide that specifically
3 hybridizes to SEQ ID NO:1 under stringent conditions.

1 11. The transgenic plant of claim 10, wherein the plant promoter is from
2 a *NBP46* gene.

1 12. The transgenic plant of claim 11, wherein the *NBP46* gene is as
2 shown in SEQ ID NO:3.

1 13. The transgenic plant of claim 10, wherein the heterologous *NBP46*
2 polynucleotide encodes a NBP46 polypeptide.

1 14. The transgenic plant of claim 13, wherein the NBP46 polypeptide is
2 SEQ ID NO:2.

1 15. The transgenic plant of claim 10, which is not a legume.

1 16. A method of modulating rhizobial interaction in a plant, the method
2 comprising introducing into the plant an expression cassette containing a plant promoter
3 operably linked to a heterologous *NBP46* polynucleotide that specifically hybridizes to
4 SEQ ID NO:1 under stringent conditions.

1 17. The method of claim 16, wherein the heterologous *NBP46*
2 polynucleotide is SEQ ID NO:1.

1 18. The method of claim 16, wherein the plant promoter is from a
2 *NBP46* gene.

1 19. The method of claim 16, wherein the heterologous *NBP46*
2 polynucleotide encodes a NBP46 polypeptide.

- 1 20. The method of claim 19, wherein the NBP46 polypeptide has an
2 amino acid sequence as shown in SEQ ID NO:2.
- 1 21. The method of claim 16, wherein the plant is not a legume.
- 1 22. The method of claim 16, wherein the expression cassette is
2 introduced into the plant through a sexual cross.
- 1 23. A method of modulating phosphohydrolase activity in a plant, the
2 method comprising introducing into the plant an expression cassette containing a plant
3 promoter operably linked to a heterologous *NBP46* polynucleotide that specifically
4 hybridizes to SEQ ID NO:1 under stringent conditions.
- 1 24. The method of claim 23, wherein the heterologous *NBP46*
2 polynucleotide is SEQ ID NO:1.
- 1 25. The method of claim 23, wherein the plant promoter is from a
2 *NBP46* gene.
- 1 26. The method of claim 23, wherein the heterologous *NBP46*
2 polynucleotide encodes a NBP46 polypeptide.
- 1 27. The method of claim 26, wherein the NBP46 polypeptide has an
2 amino acid sequence as shown in SEQ ID NO:2.
- 1 28. The method of claim 23, wherein the plant is not a legume.
- 1 29. The method of claim 23, wherein the expression cassette is
2 introduced into the plant through a sexual cross.

1/4

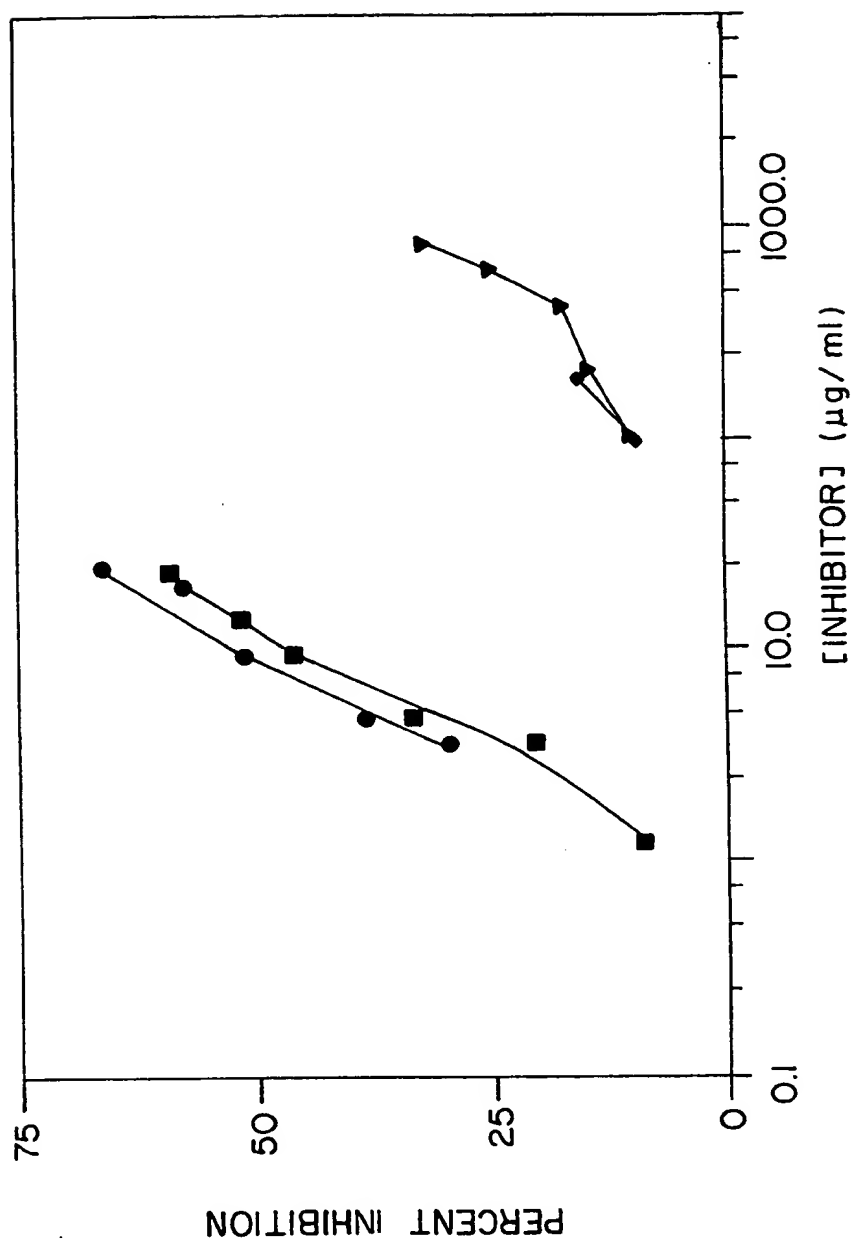


FIG. 1A.

2/4

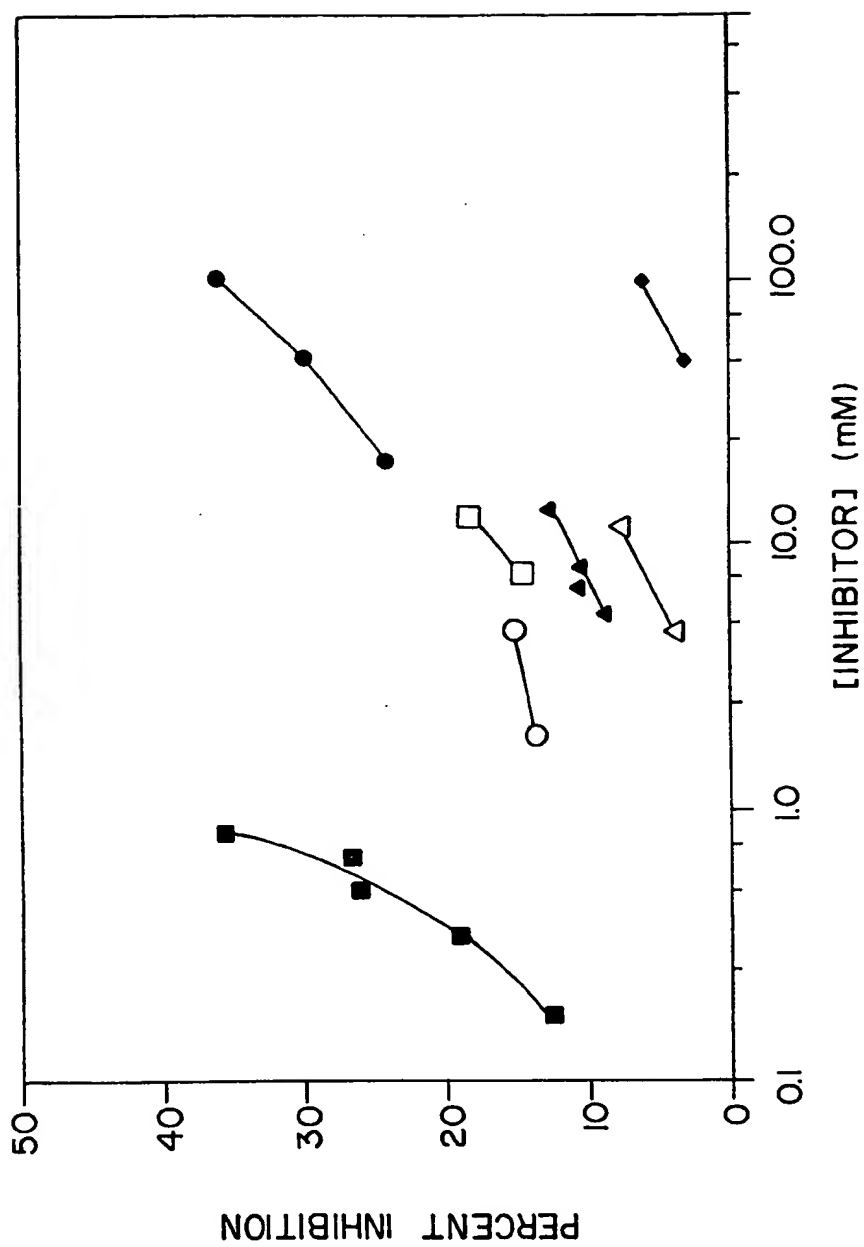


FIG. 1B.

3/4

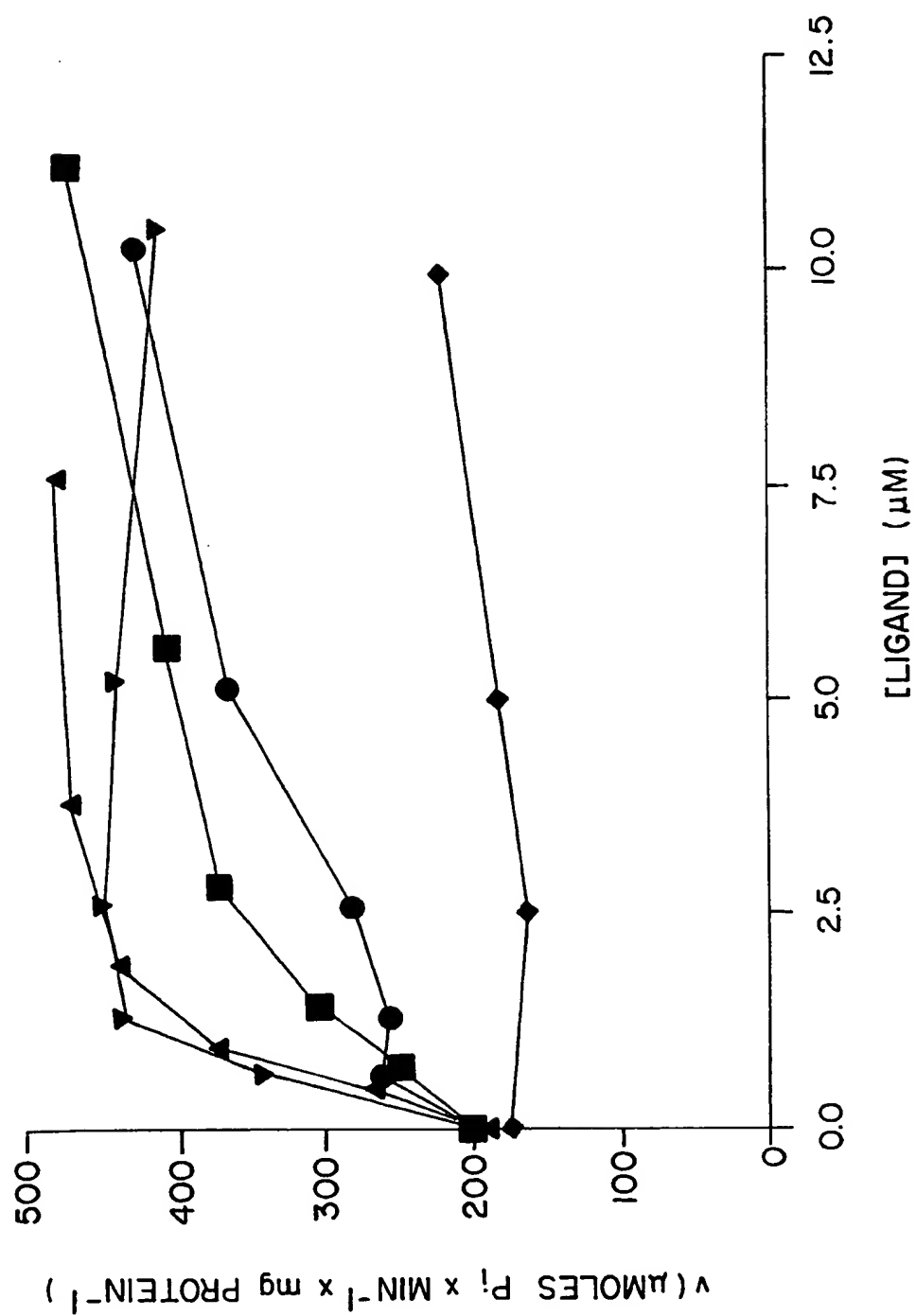


FIG. 2.

4/4

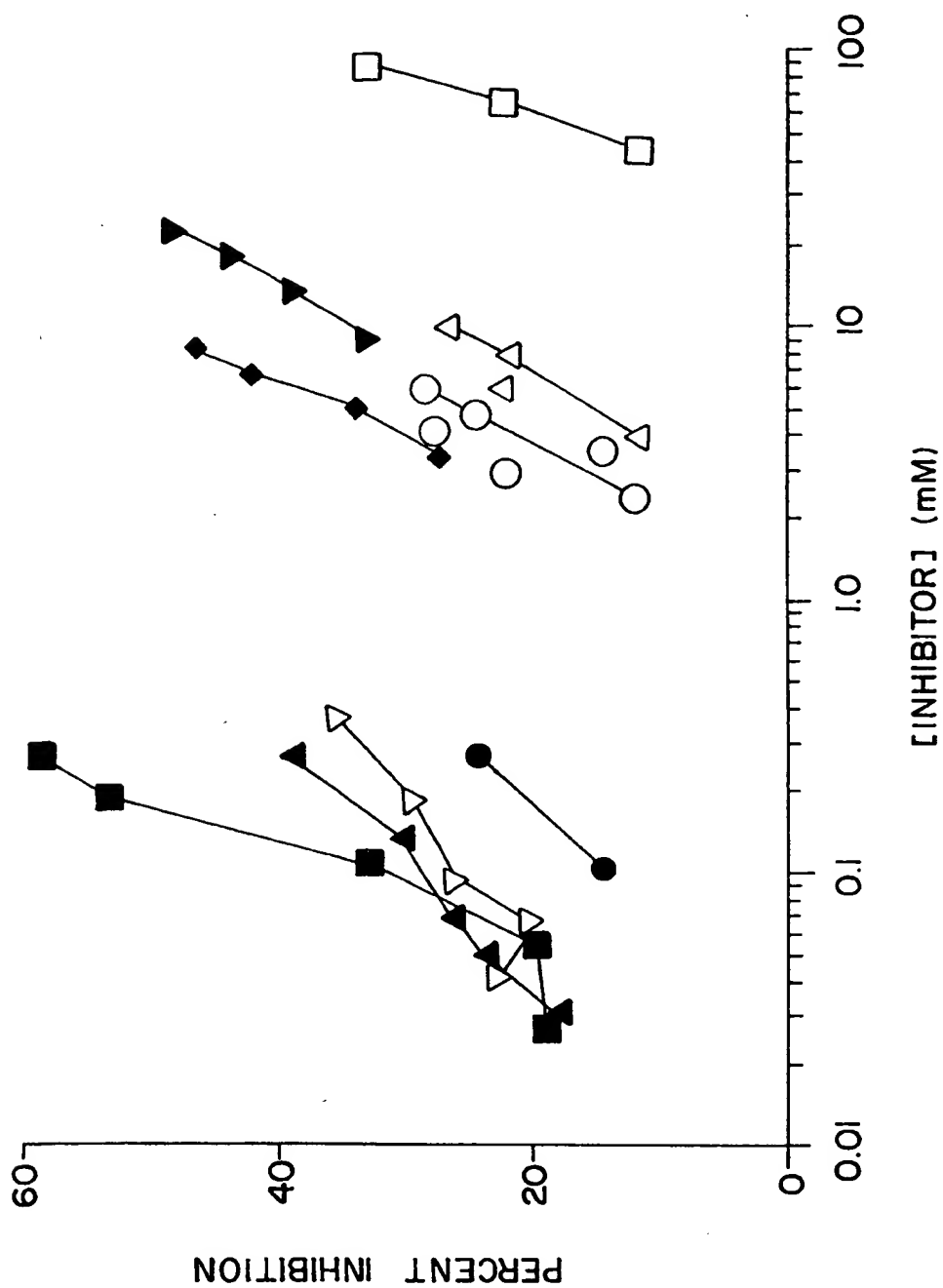


FIG. 3.

SEQ ID NO: 1 Complete cDNA sequence of *DB46*

G A A A C T G A A A C G A G T A C T C T T T C A G T G G T G A G G T T 35
C T G A G A G A T T C A G A A A T G A A T T G G G T G T G G C C A A A 70
G A C A A A G A G C A T G A G C T T C C T A C T C C T C A T C A C T T 105
T T C T A C T C T T C T C A T T G C C A A A A C T T T C T T C T T C G 140
C A A T A T G T T G G G A A C A G T A T C T T A C T A A A T C A T C G 175
T A A G A T A C T T C C C A A C C A G G A A C T C C T T A C C T C T T 210
A C G C T G T C A T C T T T G A T G C T G G T A G C T C T G G G A G T 245
C G T G T C C A T G T C T T C A A T T T T G A C C A G A A C T T A G A 280
T C T C C T G C A C A T T G G C A A T G A C C T C G A G T T T A C A A 315
A A A A G A T C A A A C C C G G T T T G A G C T C A T A C G C T G A T 350
A A G C C T G A A A A A G C T G C A G A A T C T C T C A T T C C A C T 385
T T T G G A G G A A G C T G A A G A T G T T G T C C C T G A G G A A C 420
T G C A C C C A A G A C A C C C C T T A A G C T T G G G G C A A C A 455
G C A G G T T T G A G G C T C T T G G A T G G G G A T G C T G C T G A 490
A A A G A T A T T G C A A G C G G T T A G G G A A A T G T T C A G G A 525
A C A G A A G T T C C C T G A G C G T T C A A C C T G A T G C A G T A 560
T C T G T T A T T G A T G G A A C C C A A G A A G G T T C T T A C T T 595
A T G G G T T A C A G T T A A C T A T C T G T T A G G A A A G T T G G 630
G A A A G A A G T T T A C A A A A A C T G T G G G A G T G A T A G A T 665
C T T G G A G G T G C T T C A G T T C A A A A T G G C C C A A A A C C A C 700
C T C A A G A A A T A C A G C T A A A A A T G C C C C A A A A C C A C 735
C A C A A G G A G A G G A T C C A T A C A T G A A G A A G C T T G T A 770
C T C A A G G G A A A G A A A T A T G A C C T T T A T G T T C A C A G 805
T T A C T T G C G T T A T G G T A A C G A C G C A G C A C G T G T T A 840
A G A T T T T T A A G A C C A C T G A T G G T G C T G C T A G T C C T 875
T G T C T A T T G G C A G G C T A T G A A G A T A T A T A C A G A T A 910
T T C C G G A G A A T C G T A C A A T A T C T A T G G T C C C A C T T 945
C T G G T G C C A A C T T A A T G A G T G C C G T G A C C T A G C T 980
C T T C A G A T T C T C A G A T T G A A T G A G C C A T G T T C C C A 1015
T G A A A A C T G C A C C T T T G G T G G G A T A T G G G A T G G T G 1050
G A A A A G G A A G T G G A C A G A A A A A C C T T G T T G T T A C T 1085
T C A G C T T T C T A C T A T A G G T C T T C T G A G G T T G G T T T 1120
T G T C A C T C C T C C C A A T T C C A A A A A T C G C C C T C T G G 1155
A T T T T G A A A C T G C A G C T A A A C A A G C T T G T A G T T T A 1190
A C A T T C G A G G A A G C G A A A T C C A C T T T T C C A A A T G T 1225
T G A G A A A G A T A A A C T T T C C A T T T G T A T G C G T G G A T T 1260
T C A C A T A C C A G T A T A C A T T G C T T G T T G A T G G A T T T 1295
G G C C T A G A T C C A G A G C A A G A G A T T A C A G T G G C A G A 1330
A G G A A T T G A A T A T C A A G A T G C C A T T G T G G A A A C A G 1365
C A T G G C C T C T A G G A A C T G C C A T A G A A G C C A T A T C A 1400
T C T T T G C C T A A A T T T A A T C G T C T A A T G T A T T T T A T 1435
C T A A G C C A T G T C C T C C A C T T A T G A C C A C T T T A A T T 1470
A A A A T A A A A C T C A C C C T T T T C A C T A A A A A A A A A A 1505
A A A A A A A A G T C C T T T T T A T T C C A T T G A G T A T C A A 1540
G T G T T A A T T T G T T T C T G A C A A A T G G A G G T G T A A A 1575
G T G A A A C A A A G T A T G T T T T G T C A G A T A C G A A T G G 1610
A A G T A G G G T T A T G A T G A A A A A A A A A A A A A A A A 1643

SEQ ID NO: 2

M N W V W P K T K S M S F L L L I T F L L F S L P K L S S S Q Y V G N S I L L N -9
K R K I L P N Q E L L L T S S Y A V I F D A G S S G S R V H V F N F D Q N L D L L H 32
I G N D L E P T K K I K P G L S S Y A D K P R K A A E S L I P H E E A E D V V 72
P E E L H P K T P L K L G A T A G L R L L D G D A A E K I L Q A V R E M F R N R 112
S S L S V Q P D A V S V I D G T Q E G S Y L W V T V N Y L L G K L G K K F T K T 152
V G V I D L G G A S V Q M A Y A V S R N T A K N A P K P P Q G E D P Y M K K L V 192
L K G K K Y D L Y V H S Y L R Y G N D A A R V K I F K T T D G A A S P C L L A G 232
Y E D I Y R Y S G E S Y N I Y G P T S G A N F N E C R D L A L Q I L R L N B P C 272
S H E N C T F G G I N D G G K G S G Q K N L V V T S A F Y Y R S S E V G P V T P 312
P N S K N R P L D F E T A A K Q A C S L T F E E A K S T P P N V E K D K L P F V 352
C V D F T Y Q Y T L L V D G F G L D P P E Q E I T V A E G I E Y Q D A I V E T A W 392
P L G T A I R A I S S L P K F N R L M Y F I 414

SEQ ID NO: 3

Genomic DNA sequence of Dolichos biflorus NBP46
(Bold, underlined segments indicate exons)

```

CTAGATGTGA AGTGATTTTA ATCTTGCAAC TGGTGTAAT AAATCATAAT ACAATATCTT    60
ATCTTAAAAA TAAAATCTTC ATAAAAAATA AATATAATGA TTAAATTATC ATAAATAAAT    120
AAGTAATTAT TTCCTTACCT AACATGATGG CCAGCTCATA TAATAACATC GCTTCTTGGA    180
GCATATCAAT GACGAAAACG TGGACGCAAA TTATTGGCCT CGGGGATCTG CTTTCTGCAA    240
ATACTTGTTT CTCCCAGAA CCGGATTCTC ATTAATTCT AGTTGTTCTC GTAAATTGCT    300
CACTTTATTT TCATTGTAAA GTAAAAATAA TTTTCTACTA AAAACGATAT TCACCATGTT    360
AGTCACATAC ACATTCAATA ATATTTAAAA TGTTATTTAT TTAATGGGAA GAAGATTTAA    420
TAATTGGGGT TAGTCTTAC AATAATACAT ACTCAACAAA ATTTTTCCTC AAATATCACA    480
CGATAGTAAT ATATTAATCT AATATAATCT CACAAAATCA TCTCCATATT TATATATTC    540
ATATAGATGA TGTATCATG GACGTGGATC TCTGCGACCA TAGCATTTTA CATCTATATA    600
TAGTGGCAAG AGTGACGATT AGTGCAAACT GAAACGAGTA CTCTTTCAGT GGTGAGGTTT    660
TGAGAGATTG AGAAATGAAT TGGGTGTGGC CAAAGACAAA GAGCATGAGC TTCCTACTCC    720
TCATCACTTT TCTACTCTTC TCATTGCCAA AACTTCTTTC TTCGCAATAT GTTGGGAACA    780
GTATCTTACT AAATCATCGT AAGATACTTC CCAACCAGGA ACTCCTTACC TCTTACGCTG    840
TCATCTTTGA TGCTGGTAGC TCTGGGAGTC GTGTCCATGT CTTCAATTTT GACCAGAACT    900
TAGATCTCCT GCACATTGGC AATGACCTCG AGTTTACAAA AAAGGTCAAA CTGAAACCTT    960
AAATTATTCA TTATTATTTT CTTTCATCTTA CTCTTACATT CTTCTTCATT ATTCTGGTGC    1020
AGATCAAACC CGGTTTGAGC TCATACGCTG ATAAGCCTGA AAAAGCTGCA GAATCTCTCA    1080
TTCCACTTTT GGAGGAAGCT GAAGATGTTG TCCCTGAGGA ACTGCACCCC AAGACACCCC    1140
TTAAGCTTGG GGTGAGTATT TCTCATCTCT ACTTTTGCCA CAGATTAATA TGTCACACTT    1200
TTACATGAAA CATGATTAAG TTCTTTAAAC ATGTTGATTA AAGGGTGACA GTTTGTATTT    1260
TTTAATCAAG TAATCTAGAA CTTAAACTAT GGTAATAATA TAAAATGAAT ATGAAACTAA    1320
TATATTCTGA TGGAACAGAA GAAAGCAATA TCAAGAGAGA CAAAACACAC ACTTTGATGA    1380
GCTCTATCTT TTAAACAAAA AATGGAATTG AAAGACCAAA TAAATAGGC ATTAGCCCAT    1440
ATCATAAAAA CTTTGTAAA ATATTAATAG AAAGTAAATG AACACTATAT ATGATGCATA    1500
CGTAGAAAAA GTAAAAGGAT TTTTGAGATA ATATCTTTTG ATGTTGAATG TGAATGCAGG    1560
CAACAGCAGG TTTGAGGCTC TTGGATGGGG ATGCTGCTGA AAAGATATTG CAAGCGGTAA    1620
CCATGAGCTT AGTTCATTTT CTTATGTTAT TAACTACGCT TTCAATGTCT TAACTTTCGT    1680

```

TTTCTCTCAT	GTTGAAGGTT	AGGGAAATGT	TCAGGAACAG	AAGTTCCTG	AGCGTTCAAC	1740
<u>CTGATGCAGT</u>	<u>ATCTGTTATT</u>	<u>GATGGAACCC</u>	<u>AAGAAGGTTT</u>	<u>TTACTTATGG</u>	GTATGACTTA	1800
CTTAAAGTTT	ATTTTATCA	GAATTCATTC	TAATTTTTTT	ACTTAAGAAG	ATGGAAGAAA	1860
GTGTGATCAC	ATTACCTAGG	ACATTCATCT	TATTTAAAAT	AATTTATTGC	AAAATAATAC	1920
TATTTTTTAA	TTAGAATTGA	TATTTGCGTA	TATTGTGAAA	AAGAAAAGTA	GATTGATTTT	1980
TCATTATGGT	AAAGTATTTT	AATAAATTTT	TATTAACCTCT	TTTTTAACTT	TAAAAAATAT	2040
AGGATCACTT	TATGTGTGTG	GTGACATGCC	ATACCCCATATA	TGGACAATTA	CTGACATGCC	2100
ATACCCGATA	TATTAATATT	TTATCAATTG	TCAATTTATT	TATTGTAACCT	ACTTTAAAAA	2160
ATACTTTTAA	TTAAATCATT	GAGGTATCGC	TTTAGTTTTT	TTTTAAATTC	GAAAAAATAA	2220
TTAATAATTA	TTAGATATAC	TGGAAGAATT	TCCGAAGGAT	ATTCATATCC	ATATATATCT	2280
TGTAAGATAA	CATTTTTTAT	TGAACAAATG	CAACACTATC	TCTAAATATG	ATTTTTTTAT	2340
TTATGTCGAA	TGAATCACGA	CGATATAATT	TTGTATAAGT	AATTAAATTC	ACTATTCATT	2400
TTTATTTGTT	GTGTTCTTTT	TAGGGTCCGC	CAATTAGCTA	AATCTTACCT	AAAAAGATTG	2460
CAAACAAAGA	AAAAAGAAAG	AAGCAATGAT	GAAATTAAAA	GTGGATCAAA	CCATGAGGAT	2520
ATGTTTCAAA	AAGAAGAATT	AGGTTCTTTG	TTATGTTTTT	AAAACTAGT	AGTTGGAATT	2580
TCTTAAATTC	AATTATAATT	ATTAAATAAA	ATTGTCGCT	TAATTGATAA	TATAAAATAG	2640
CATAACTGAT	ACATTTATAA	ATTATATTTT	ATATTAAAAAT	TTATTTTTAT	TTTATAGATA	2700
AAATGTATTT	GGTAATATTT	ATAATATAGT	TTTAAATTAA	TTTCAAATTC	GTTGTGATCT	2760
TACTTATAAA	TTAATTATTT	TTTTCAGTTT	TCAATTATTG	CATTTTTCTT	ATAATATTCA	2820
CTATATTAAT	ATTTGACAAT	ATTTCAAAAC	ATTTTCAATA	AAAAAAAAAA	AAAAAGAAGT	2880
TCAGTAAACT	TCATATCTGC	ATTATGTTTA	TTTGAATAGT	AAAACACTAT	AAAATATATC	2940
TAATGTAAAG	GATAAACATG	CAGAGTAGTA	AAAACTTAT	TTAGAATATA	GTCATTTAAT	3000
TTTTCTTATG	ATATATCTTG	GGAATTTTGT	GTAGGTTACA	<u>GTAACTATC</u>	<u>TGTTAGGAAA</u>	3060
<u>GTTGGGAAAG</u>	<u>AAGTTTACAA</u>	<u>AAACTGTGGG</u>	<u>AGTGATAGAT</u>	<u>CTTGGAGGTG</u>	<u>CTTCAGTTCA</u>	3120
<u>AATGGCTTAT</u>	<u>GCTGTCTCAA</u>	<u>GAAATACAGC</u>	<u>TAAAAATGCC</u>	<u>CCAAAACCAC</u>	<u>CACAAGGAGA</u>	3180
<u>GGATCCATAC</u>	<u>ATGAAGAAGC</u>	<u>TTGTACTCAA</u>	<u>GGGAAAGAAA</u>	<u>TATGACCTTT</u>	<u>ATGTTACACAG</u>	3240
GTTACTTTCT	GTTATCATTC	ATATAGCAAA	GGAACAATTA	TCATTTCAAT	TTCTAAAATA	3300
TATTTATAAT	CTCTAAAATC	AAATAACATA	AAAAAATGGT	AATATAATGT	TGCGTTTTTG	3360
GATTGTTTGG	ATTAAAGGGT	AAATTTGAAG	AAGAAAAAAA	ATAATAAATA	AAGAAAAAGA	3420

GAAAAAAAT	AAGATTGTTT	GGATTATTAG	AAAGAGAAAA	AGTTGAATAA	TTATTTTTAT	3480
ATTTTAATAT	TATTTTAATT	ATTATTATT	ATGAAAATAA	AATATTTATT	TTTAAATTTA	3540
TATTTTATTA	TTATTTTTTA	ATTTTATTAT	TATAAAAATA	TAAATATTAT	TAATAATTAT	3600
TATTTTAATT	TTATTTATTA	ATATAATATA	ATAATAAATA	AAATATTAAT	ATTTTATGTT	3660
ATATTATATA	ATATTTAATT	ATACATATGT	ATTTTTTTTC	TGCAAATTTT	TACCTTTTAA	3720
GCGGAGAAGA	TGAAGGGCAT	AAATTGTTCT	CGAAATTAGT	TATATTTTGT	TCAATTTTAA	3780
CAAAATCATC	TCAAATCAGT	CTTCATAAAT	AGTATTTATG	TAGATCCAAA	TAGAGGCTTA	3840
ACGTGGTCTA	GTTGTACAAA	CCTAAAAGGT	GTTTCTTTTT	TTCTTTAATT	TGAAGAACTA	3900
GAATATTGTT	TTTCAATTTG	AAAGACGAAG	GCAAACCTTA	CCAAATTTAG	AAAAAGTAAA	3960
AACCTGGTTA	ACTTTATAAC	GAATGTCAGA	AAAAATGGTA	GGTATGTTAT	AAATACTTCT	4020
GATATCAAAA	TGGCAAAAAC	TCCAGAGTCT	CACCTCCAAG	AATCATCACT	TTTTCTCACC	4080
TTAATCTGAA	ATAATGAATG	CTTACTTTTT	TTAAGATATT	TATAGATATC	TATAATCCAT	4140
TGAAGTTCAG	TGTAGTGTA	ATAAATTATA	ATGTAAAAAC	CTATACACTG	AGTACAGATC	4200
CATGTGTAGT	TACTTTTTTA	TGGTTAACT	GATAAATATG	CATGAGTCAT	GTCATGGCTA	4260
ACGTACAGGT	CTTAATCAAC	TTCTTTGTTG	CAGTTACTTG	CGTTATGGTA	ACGACGCAGC	4320
<u>ACGTGTTAAG</u>	<u>ATTTTTAAGA</u>	<u>CCACTGATGG</u>	<u>TGCTGCTAGT</u>	<u>CCTTGCCTAT</u>	<u>TGGCAGGCTA</u>	4380
<u>TGAAGGTA</u>	<u>TAAAGTATTC</u>	<u>TTTGTACAA</u>	<u>ACCCTAATGT</u>	<u>TACTTTCTTA</u>	<u>TTCTGTCATT</u>	4440
CAGAATAGTG	CAAAGGACTG	AAACTAGAAA	GGATTCCAAT	TCACTACAAG	AAGAAAAAAA	4500
AGTAGTGATT	TAGTGACCAA	AGTTACTTTT	TCCTCACTGA	GTTCTATTGA	AATGCAGAAA	4560
CTTGTTGCAG	ATATTTTAAA	TACATATTAA	GTGTTTGTG	AGTACTGCAT	TTGTTTTTAG	4620
TGATTTCAAG	TCGAGTTTTT	TCTTGAAGCA	TTAAAGCTGC	AAATAACATG	TGGGTCTTTT	4680
TTCTATCTTT	AAAGATATAT	ACAGATATTC	CGGAGAATCG	TACAATATCT	ATGGTCCCAC	4740
<u>TTCTGGTGCC</u>	<u>AACTTTAATG</u>	<u>AGTGCCGTGA</u>	<u>CCTAGCTCTT</u>	<u>CAGATTCTCA</u>	<u>GATTGAATGA</u>	4800
<u>GCCATGTTCC</u>	<u>CATGAAAAC</u>	<u>GCACCTTTGG</u>	<u>TGGGATATGG</u>	<u>GATGGTGGAA</u>	<u>AAGGAAGTGG</u>	4860
<u>ACAGAAAAAC</u>	<u>CTTGTTGTTA</u>	<u>CTTCAGCTTT</u>	<u>CTACTATAGG</u>	<u>TCTTCTGAGG</u>	<u>TATCCATTCT</u>	4920
CTGTTAATTT	CTTGTTTACT	TTGATTACTT	ATTTGTTTTT	ATACCAATAA	ATTTTACATT	4980
ATAGTTTATA	CTGTGCTAAT	TTGTTGTTT	TTAGGTTGGT	TTTGTCCTC	CTCCCAATTC	5040
<u>CAAAAATCGC</u>	<u>CCTCTGGATT</u>	<u>TTGAACTGC</u>	<u>AGCTAAACAA</u>	<u>GCTTGTAGTT</u>	<u>TAACATTCCA</u>	5100

GGAAGCGAAA TCCACTTTTC CAAATGTTGA GAAAGATAAA CTTCCATTG TATGCGTGGA 5160
TTTCACATAC CAGTATACAT TGCTTGTTGA TGGATTTGGT ATGTTTTTCAT AATTAATTAC 5220
 CAAGTTGATA TTAACTTCT TCCAAAAAAC TATGTTTTCT TTTGTCTTCC AACACTGACT 5280
 CCTAATTCAA CTTTTGGCAG GCCTAGATCC AGAGCAAGAG ATTACAGTGG CAGAAGGAAT 5340
TGAATATCAA GATGCCATTG TGGAACAGC ATGGCCTCTA GGAAGTCCCA TAGAAGCCAT 5400
ATCATCTTTG CCTAAATTA ATCGTCTAAT GTATTTTATC TAAGCCATGT CCTCCACTTA 5460
TGACCACTTT AATTAAAATA AAAGTCACCC TTTTACTAG TCCTTTTTTA TTCCATTGAG 5520
TATCAAGTGT TAATTTGTTT CTGACAAATG GAGGTGTAAA AGTGAAACAA AGTATGTTTT 5580
TGTCAGATAC GAATGGAAGT AGGGTTATGA TGACCTGCCA ATTAACATAAT ACTCTGACTT 5640
 CTTTCATCCT CTTATTTTAA TTTTGAAAAA AAAAATCATA TATGTAATCG GGAAAATTTG 5700
 ATTTGCAACT TAAAAAATG ACCAAATAAA TTTTCCTAAA TTCCTCTGCA ACATATACAA 5760
 GGACCACAAA ATTGAATCTG TTTCTTTAAT GGAATAAGTA CTTTTTGAAA AACTATCATA 5820
 TTAGTAACT TATCTTTTC ATCTAACAGG CAGCAAAAT AATTGCATGA ACGGATCCAA 5880
 TTAATCTCT CGTACAGCTC CAGATAAGAA GCGTTTAATG AGATAAATTG TTGGATAATA 5940
 TATGTTGGGT GTGGGTGGAT TATGATACTA TCGATAATAA ATTTGGAATC TAATTAAATT 6000
 TTATAAAATT AATTATCAA TATATAATAT TTTATATATA TTAATTTGAT AATATTTTTA 6060
 ATAATTTTAT ATTTTAAATA TTTAATTTTA ATTTAAGGAA ATTTTAAAGA TAATTAATTT 6120
 TTTATTTTTA TTTTTTTGTA TAGTACTCAG GACATAATAA TGTTATTAAT TTAAATAAGA 6180
 CTTAAATATA TATTTTCTT ATAATGCTTA AATCTCAGTC TTATTATTGC TATCACATAA 6240
 TGACACGAAC TAACTAGCTT CACTC 6300

SEQ ID NO:4 5' Degenerate primer sequence
TA(T/C)GCNGTNAT(T/C)TT(T/C)GATCG

SEQ ID NO:5 3' Degenerate primer sequence
AT(A/G)TT(A/G)TA(T/A/G)AT(G/A)CCNGG

SEQ ID NO:6 3' RACE primer
CGTCCGATACTTCTATA

SEQ ID NO:7 5' RACE primer
AACTTAGATCTCCTGCAC

SEQ ID NOS: 8-9

```

5' 11      20      29      38      47      56
   CAA ATT AAG AAC ATG GAG TTC CTA ATT ACA CTC ATT GCC ACT TTT TTA CTC TTG
   ---
   Q   I   K   N   M   E   F   L   I   T   L   I   A   T   F   L   L   L

      65      74      83      92      101      110
   TTA ATG CCT GCA ATC ACT TCC TCC CAA TAT TTA GGA AAC AAC CTA CTC ACT AAT
   ---
   L   M   P   A   I   T   S   S   Q   Y   L   G   N   N   L   L   T   N

      119      128      137      146      155      164
   CGA AAG ATT TTC CAA AAA CAA GAA ACC TTA ACC TCT TAC GCT GTC ATA TTT GAT
   ---
   R   K   I   F   Q   K   Q   E   T   L   T   S   Y   A   V   I   F   D

      173      182      191      200      209      218
   GCT GGT AGC ACT GGT ACT CGT GTC CAT GTT TAC CAT TTT GAT CAG AAC TTA GAT
   ---
   A   G   S   T   G   T   R   V   H   V   Y   H   F   D   Q   N   L   D

      227      236      245      254      263      272
   CTA CTT CAC ATT GGC AAT GAT ATT GAG TTT GTT GAC AAG ATC AAA CCA GGT TTG
   ---
   L   L   H   I   G   N   D   I   E   F   V   D   K   I   K   P   G   L

      281      290      299      308      317      326
   AGT GCA TAT GGG GAT AAT CCT GAA CAA GCA GCA AAA TCT CTC ATT CCA CTT TTG
   ---
   S   A   Y   G   D   N   P   E   Q   A   A   K   S   L   I   P   L   L

      335      344      353      362      371      380
   GAG GAA GCA GAA GAT GTG GTT CCT GAG GAT CTG CAC CCC AAA ACA CCC CTT AGG
   ---
   E   E   A   E   D   V   V   P   E   D   L   H   P   K   T   P   L   R

      389      398      407      416      425      434
   CTT GGG GCA ACC GCA GGT TTG AGG CTT TTG AAT GGG GAT GCT GCT GAA AAG ATA
   ---
   L   G   A   T   A   G   L   R   L   L   N   G   D   A   A   E   K   I

      443      452      461      470      479      488
   TTG CAA GCG ACA AGG AAT ATG TTC AGC AAC AGA AGT ACC CTC AAC GTT CAA CGT
   ---
   L   Q   A   T   R   N   M   F   S   N   R   S   T   L   N   V   Q   R

      497      506      515      524      533      542
   GAT GCA GTT TCT ATT ATT GAT GGA ACC CAA GAA GGT TCT TAT ATG TGG GTG ACA
   ---
   D   A   V   S   I   I   D   G   T   Q   E   G   S   Y   M   W   V   T

      551      560      569      578      587      596
   GTT AAC TAT GTA TTG GGG AAT TTG GGA AAA AGC TTC ACA AAA TCA GTG GGA GTA
   ---
   V   N   Y   V   L   G   N   L   G   K   S   F   T   K   S   V   G   V

      605      614      623      632      641      650
   ATT GAC CTT GGA GGT GGT TCA GTT CAA ATG ACA TAT GCA GTG TCA AAG AAA ACA
   ---
   I   D   L   G   G   G   S   V   Q   M   T   Y   A   V   S   K   K   T

```

659	668	677	686	695	704
GCA AAA AAT GCT CCT	AAA GTT GCT	GAT GGA GAG GAT	CCA TAT ATT AAG	AAG CTT	
A K N A P K V A D G E D P Y I K					K L
713	722	731	740	749	758
GTG CTC AAG GGA AAG	CAA TAT GAT	CTC TAT GTT CAT	AGT TAC TTG CGT	TTT GGC	
V L K G K Q Y D L Y V H S Y L R					F G
767	776	785	794	803	812
AAA GAA GCA ACT CGA	GCA CAG GTT TTG	AAT GCA ACT AAT	GGA TCT GCT	AAC CCT	
K E A T R A Q V L N A T N G S A					N P
821	830	839	848	857	866
TGC ATT TTA CCT GGA	TTT AAT GGG ACC	TTT ACA TAT TCA	GGA GTG GAG	TAT AAG	
C I L P G F N G T F T Y S G V E					Y K
875	884	893	902	911	920
GCT TTT TCC CCT TCT	TCT GGC TCC AAC	TTT GAT GAT TGC	AAA GAA ATA	ATT CTT	
A F S P S S G S N F D D C K E I					I L
929	938	947	956	965	974
AAG GTT CTT AAA GTA	AAT GAT CCA TGT	CCC TAT CCG AGT	TGC ACT TTT	GGT GGA	
K V L K V N D P C P Y P S C T F					G G
983	992	1001	1010	1019	1028
ATA TGG AAT GGT GGA	GGA GGG AGT GGA	CAA AAA AAA CTT	TTT GTT ACT	TCA GCT	
I W N G G G G S G Q K K L F V T					S A
1037	1046	1055	1064	1073	1082
TTC GCT TAC CTG GCT	GAA GAT GTT GGT	ATG GTT GAG CCA	AAT AAA CCT	AAT TCC	
F A Y L A E D V G M V E P N K P					N S
1091	1100	1109	1118	1127	1136
ATA CTT CAT CCA GTA	GAT TTC GAA ATT	GAA GCT AAG CGA	GCT TGT GCA	TTA AAC	
I L H P V D F E I E A K R A C A					L N
1145	1154	1163	1172	1181	1190
TTT GAG GAT GTC AAA	TCC ACT TAT CCT	CGA CTT ACG GAT	GCA AAA CGT	CCA TAT	
F E D V K S T Y P R L T D A K R					P Y
1199	1208	1217	1226	1235	1244
GTA TGC ATG GAT CTC	TTA TAC CAA CAT	GTG TTG CTT GTT	CAT GGA TTT	GGC TTA	
V C M D L L Y Q H V L L V H G F					G L
1253	1262	1271	1280	1289	1298
GGT CCA CGA AAA GAG	ATT ACA GTA GGT	GAG GGA ATT CAA	TAT CAG AAT	TCT GTT	
G P R K E I T V G E G I Q Y Q N					S V
1307	1316	1325	1334	1343	1352
GTG GAA GCT GCA TGG	CCT CTA GGT	ACT GCC GTG	GAA GCC ATA	TCA GCG	TTA CCT
V E A A W P L G T A V E A I S A					L P

		1361		1370		1379		1388		1397		1406					
AAG	TTT	AAG	CGA	TTA	ATG	TAT	TTT	ATT	TAA	GCT	TTT	AGA	GAT	GTC	AAG	ATA	TTT
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
K	F	K	R	L	M	Y	F	I	*	A	F	R	D	V	K	I	F
		1415		1424		1433		1442		1451		1460					
CAG	TAA	CAG	CTA	ACT	TTA	TCA	AAA	ATT	AAA	TAA	AAC	TGG	CGC	ATT	TTG	TCT	TTC 3'
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Q	*	Q	L	T	L	S	K	I	K	*	N	W	R	I	L	S	F

SEQ ID NOS: 10-11

```

5'      9      18      27      36      45      54
AAG TGC TCT TCT CTC TGT AGT TAG TTG CAT TGG ACT AAA GCC ATG GAC TTC TTA
---
K   C   S   S   L   C   S   *   L   H   W   T   K   A   M   D   F   L

      63      72      81      90      99      108
ATT AGT CTC ATG ACC TTT GTG TTC ATG TTA ATG CCT GCT ATC TCT TCC TCC CAA
---
I   S   L   M   T   F   V   F   M   L   M   P   A   I   S   S   S   Q

      117     126     135     144     153     162
TAT CTC GGA AAC AAC ATT CTC ATG AAT CGT AAG ATA TTA CTC CCC AAA AAT CAG
---
Y   L   G   N   N   I   L   M   N   R   K   I   L   L   P   K   N   Q

      171     180     189     198     207     216
GAA CCA GTT ACA TCA TAC GCT GTT ATA TTT GAT GCT GGT AGC ACT GGA AGC AGA
---
E   P   V   T   S   Y   A   V   I   F   D   A   G   S   T   G   S   R

      225     234     243     252     261     270
GTC CAT GTC TAC AAT TTT GAT CAG AAC TTA GAT CTC CTT CCC GTT GAA AAC GAA
---
V   H   V   Y   N   F   D   Q   N   L   D   L   L   P   V   E   N   E

      279     288     297     306     315     324
CTT GAG TTT TAT GAT TCG GTT AAA CCC GGT TTG AGT TCA TAC GCT GGT AAT CCT
---
L   E   F   Y   D   S   V   K   P   G   L   S   S   Y   A   A   N   P

      333     342     351     360     369     378
GAA GAA GCT GCA GAA TCT CTG ATT CCA CTT CTA AAA GAA GCA GAA AAT GTG GTT
---
E   E   A   A   E   S   L   I   P   L   L   K   E   A   E   N   V   V

      387     396     405     414     423     432
CCT GTG AGC CAG CAA CCC AAC ACA CCC GTT AAG CTT GGG GCA ACT GCA GGT TTA
---
P   V   S   Q   Q   P   N   T   P   V   K   L   G   A   T   A   G   L

      441     450     459     468     477     486
AGG CTT TTG GAG GGG AAT GCT GCT GAA AAT ATA TTG CAA GCG GTC AGG GAT ATG
---
R   L   L   E   G   N   A   A   E   N   I   L   Q   A   V   R   D   M

      495     504     513     522     531     540
CTC AGC AAC AGA AGT GCC CTT AAT GTT CAA TCA GAT GCA GTA TCT ATT CTT GAT
---
L   S   N   R   S   A   L   N   V   Q   S   D   A   V   S   I   L   D

      549     558     567     576     585     594
GGA ACC CAA GAA GGT TCT TAT CTT TGG GTG ACA ATT AAC TAT CTC TTG GGG AAG
---
G   T   Q   E   G   S   Y   L   W   V   T   I   N   Y   L   L   G   K

      603     612     621     630     639     648
TTG GGA AAA AGA TTT ACA AAG ACA GTG GGA GTA GTT GAT CTA GGA GGT GGG TCA
---
L   G   K   R   F   T   K   T   V   G   V   V   D   L   G   G   G   S

```

657	666	675	684	693	
GTG CAA ATG ACA TAT	GCA GTC TCA AGG AAC ACA GCT AAA AAT	GCT CCA AAA			
V Q M T Y A V S R N T A K N A P K					
711	720	729	738	747	
CCT GAA GGA GAG GAT	CCA TAC ATA AAG AAG CTT GTA CTC CAG GGA AAG AAA				
P E G E D P Y I K K L V L Q G K K					
765	774	783	792	801	8
GAC CTT TAT GTT CAC AGT TAC TTG CGC TAT GGA AGA GAA GCA TTT CGT GCA					
D L Y V H S Y L R Y G R E A F R A I					
819	828	837	846	855	86
ATT TTC AAG GTC GCT GGT GGT TCT GCT AAT CCT TGC ATT TTA GCT GGC TTT GA					
I F K V A G G S A N P C I L A G F D					
873	882	891	900	909	918
GGG GCA TAT ACA TAT TCC GGA GCA GAG TAT AAG GTC TCG GCC CCA GCT TCA GGA					
G A Y T Y S G A E Y K V S A P A S G					
927	936	945	954	963	972
TCT AAC TTG AAT CAA TGC AGA AAG ATA GCT CTT AAG GCT CTT AAA GTG AAT GCA					
S N L N Q C R K I A L K A L K V N A					
981	990	999	1008	1017	1026
CCT TGT CCC TAT CAG AAT TGC ACT TTT GGT GGG ATA TGG AAT GGT GGA GGT GGA					
P C P Y Q N C T F G G I W N G G G G					
1035	1044	1053	1062	1071	1080
AGT GGT CAA AAA AAT CTT TTC CTT ACT TCA TCT TTC TAT TAC CTC TCT GAA GAT					
S G Q K N L F L T S S F Y Y L S E D					
1089	1098	1107	1116	1125	1134
GTT GGG ATC TTT GTG AAT AAA CCC AAT GCC AAA ATT CGT CCA GTT GAT TTG AAG					
V G I F V N K P N A K I R P V D L K					
1143	1152	1161	1170	1179	1188
ACT GCA GCT AAA CTA GCT TGT AAA ACA AAT CTT GAG GAT GCA AAA TCC AAA TAC					
T A A K L A C K T N L E D A K S K Y					
1197	1206	1215	1224	1233	1242
CCA GAT CTT TAT GAG AAA GAC AGT GTT GAA TAT GTG TGC TTG GAT CTT GTC TAC					
P D L Y E K D S V E Y V C L D L V Y					
1251	1260	1269	1278	1287	1296
GTG TAC ACA TTG CTT GTT GAT GGA TTT GGT CTT GAT CCA TTT CAA GAG GTT ACA					
V Y T L L V D G F G L D P F Q E V T					
1305	1314	1323	1332	1341	1350
GTG GCG AAT GAA ATT GAA TAT CAG GAT GCT CTT GTG GAA GCC GCA TGG CCT CTA					
V A N E I E Y Q D A L V E A A W P L					

1359	1368	1377	1386	1395	1404
GGC ACT GCC ATA GAA GCA ATA TCA TCA TTG CCT AAA TTT GAG AGA TTA ATG TAT					
---	---	---	---	---	---
G T A I E A I S S L P K F E R L M Y					
1413	1422	1431	1440	1449	1458
TTT ATT TAA ACT ACT AGT ACC TGC TTA AGC CTG GAT TAC CTG AAG AAA TAA AAT					
---	---	---	---	---	---
F I * T T S T C L S L D Y L K K * N					
1467	1476	1485			
GAA ATA AAA GCC GCA TCT TTC TTC CTT GCT T 3'					
---	---	---	---	---	---
E I K A A S F F L A					

SEQ ID NOS: 12-13

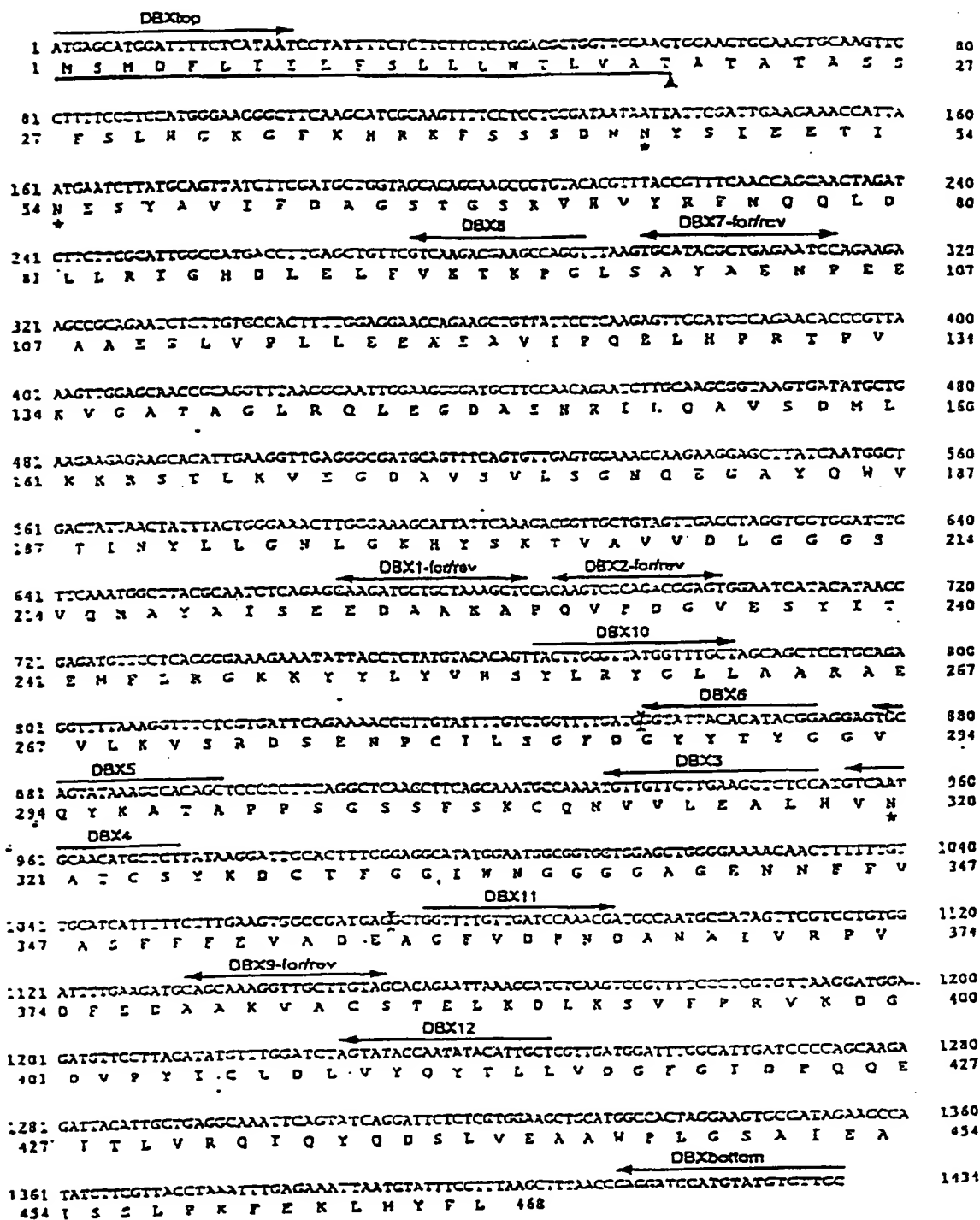


Figure 5 - DBX cDNA and deduced amino acid sequences. DBX primers are denoted by arrows. The predicted signal sequence is underlined. I - splice site, A - predicted signal sequence cleavage site, * - potential N-glycosylation site.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/16261

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :536/24.32; 800/295

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/24.32; 800/295

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAPLUS, MEDLINE, BIOSIS, AGRICOLA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	RELIC et al. Nod Factors of Rhizobium are a Key to the Legume Door. Mol Microbiol. July 1994, Vol. 13, No. 1, pages 171-178, see entire documentation.	1-29
Y	BAUER et al. Alfalfa Enod12 Genes are Differentially Regulated During Nodule Development by Nod Factors and Rhizobium Invasion. Plant Physiol. June 1994, Vol. 105, No. 2, pages 585-592, see entire documentation.	1-29
Y	QUINN et al. Isolation and Characterization of a Lectin from the Roots of Dolichos Biflorus. Arch Biochem Biophys. 01 November 1987, Vol. 258, No. 2, pages 535-544, see entire documentation.	1-29

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

26 OCTOBER 1998

Date of mailing of the international search report

12 NOV 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

OUSAMA M-FAIZ ZAGHMOUT

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/16261

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

A01N 43/16; A61K 31/715; C07H 3/06; C12P 19/26; C12N 5/04; A01H 4/00